Chapter 7

FPGA-Based Hardware

Besides off-the-shelf graphics cards, reconfigurable hardware has been gaining attention in the field of massively parallel high-performance computing as well. Several computer manufacturers offer FPGA-based accelerator components together with dedicated libraries in order to speed up the execution of numerically intensive codes. Such hardware components are particularly appropriate for applications in the signal processing domain such as image or video compression, for instance.

Generally speaking, an FPGA is a dynamically reconfigurable microchip that covers logical hardware blocks (e.g., look-up tables), arithmetic units (e.g., multiply-add blocks), as well as I/O functionality [Meye08]. Today's FPGA designs typically run at clock rates of 100 MHz up to 500 MHz. An FPGA must be loaded with an appropriate firmware (also known as bitstream) before it provides its functionality. The generation of such a bitstream is commonly a complex process that involves a series of software tools covering the steps of netlist synthesis and place-and-route, amongst others. However, the details of the firmware generation process are far beyond the scope of this thesis. In short, a higher-level representation of the hardware functionality (typically given as VHDL code or as Verilog code) is turned into the actual firmware to be loaded into the FPGA.

The main contributions of this chapter have been presented in part at the Symposium on Simulation Techniques 2005 [Sche05].

7.1 Architecture of the ImageProX Hardware

As an example of an FPGA-based accelerator hardware, we focus on the ImageProX (image processing accelerator) board that has been developed at Siemens Healthcare and was released in 2006. The ImageProX board uses either a 64 bit PCI interface (66 MHz) or a 64 bit PCI-X interface (133 MHz) to connect to the host PC. It covers nine Xilinx Virtex-4 FPGAs (1× Virtex-4 SX55, 8× Virtex-4 SX35), each of which is equipped with up to 1 GB of external DDR2 SDRAM memory. The ImageProX board comprises two identically organized rings of four Virtex-4 SX35 chips each, with the even more powerful Virtex-4 SX55 FPGA representing the core control and interface unit of the design.

Figure 7.1 clearly shows the ImageProX architecture with its nine Xilinx FPGAs. See [Heig07] for further architecture details.
Assuming a system clock rate of 200 MHz and counting the aforementioned arithmetic units of the FPGAs only (while ignoring that standard FPGA logic can also be used for computing purposes), ImageProX offers a peak performance of more than 800 Giga operations per second. It must be pointed out that FPGAs currently offer fixed-point arithmetic only. If, however, floating-point arithmetic is required for the sake of numerical accuracy, floating-point units must be built "manually" using the available fixed-point units. This typically leads to inefficient designs that cannot cope with compute architectures that natively support floating-point operations.

7.2 Feldkamp Algorithm

Our ImageProX-based implementation of the FDK method covers the two essential steps of filtering the rows of each individual projection as well as back-projecting the filtered projection data into the volume.

7.2.1 Filtering

As was already mentioned above, a significant downside to our FPGA-based approach is that it is no longer possible to use floating-point arithmetic, as is the case for conventional CPU-based implementations, for example. Consequently, it is necessary to realize the whole processing chain of the convolution using fixed-point numbers instead. In order to simulate the effects of fixed-point calculations on the accuracy of the final numerical results, we have developed a highly flexible and bit-accurate software prototype of the hardware design. It covers both the FFT routines as well as different scaling strategies of the involved fixed-point data types. Additionally, various hardware restrictions of the FPGA architecture had to be considered throughout the simulation task to achieve optimal performance and to meet the resource restrictions on the chips.